

Inferring the regulatory network behind a gene expression experiment

Marta Bleda^{1,2}, Ignacio Medina¹, Roberto Alonso¹, Alejandro De María¹, Francisco Salavert^{1,2} and Joaquín Dopazo^{1,2,3,*}

¹Department of Bioinformatics and Genomics, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain,

²CIBER de Enfermedades Raras (CIBERER), Valencia, 46012, Spain and ³Functional Genomics Node (INB) at CIPF, Valencia, 46012, Spain

Received February 28, 2012; Revised April 30, 2012; Accepted May 21, 2012

ABSTRACT

Transcription factors (TFs) and miRNAs are the most important dynamic regulators in the control of gene expression in multicellular organisms. These regulatory elements play crucial roles in development, cell cycling and cell signaling, and they have also been associated with many diseases. The Regulatory Network Analysis Tool (RENATO) web server makes the exploration of regulatory networks easy, enabling a better understanding of functional modularity and network integrity under specific perturbations. RENATO is suitable for the analysis of the result of expression profiling experiments. The program analyses lists of genes and search for the regulators compatible with its activation or deactivation. Tests of single enrichment or gene set enrichment allow the selection of the subset of TFs or miRNAs significantly involved in the regulation of the query genes. RENATO also offers an interactive advanced graphical interface that allows exploring the regulatory network found. RENATO is available at: <http://renato.bioinfo.cipf.es/>.

INTRODUCTION

Understanding the regulatory mechanisms that explain a gene expression profiling experiment remains a difficult task yet. Transcription factors (TFs) and miRNAs play a crucial role in the dynamic regulation of the network of gene expression (1). Alterations in these elements have been extensively related with human malignancies, including development diseases (2) and cancer (3,4). Here, we present Regulatory Network Analysis Tool (RENATO), a network-based analysis web tool, for the

interpretation and visualization of transcriptional regulatory information. RENATO contains information on regulatory elements in genes, such as transcription factor binding sites and miRNA complementarity targets. RENATO has been designed to identify common regulatory elements in a list of genes. It maps such genes to the regulatory network, extracts the corresponding regulatory connections and evaluates each regulator for significant over-representation (5) of targets in the list. Ranked gene lists can also be analysed with RENATO by using a strategy similar to the gene set enrichment analysis (GSEA) (6). Previous knowledge of the relationship between diseases and the deregulation of these elements is also included. To make these resources easily accessible, we have also implemented a set of RESTful application programming interface web services, where regulatory information can be easily retrieved.

IMPLEMENTATION

RENATO inputs a list of gene identifiers, typically (but not restricted to) differentially expressed genes in transcriptomic profiling experiments (microarray or RNA-seq). The output consists of a list of regulatory elements of the genes in the list along with the corresponding *P* value of enrichment. Results are represented graphically through an interactive user-friendly web interface that relates regulatory elements to their target genes.

Figure 1 shows a schematic representation of the application. A RESTful web service queries the database containing information on the regulatory elements contained in the input file. Then a test is conducted that returns the TFs or miRNAs corresponding to the enriched regulatory elements. The web application displays the results. The interface is based in modern HTML5 technologies and allows the user to explore the regulatory network found as well as executing several operations, including different changes in the format

*To whom correspondence should be addressed. Tel: +34 96 328 96 80; Fax: +34 96 328 97 01; Email: jdopazo@cipf.es

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

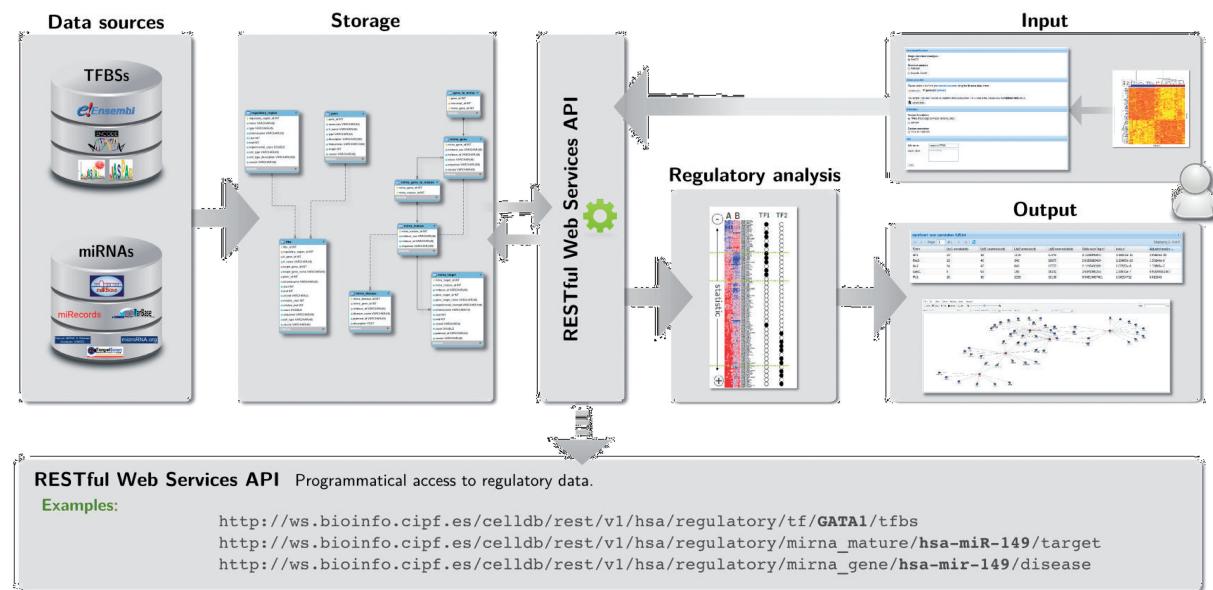


Figure 1. Schematic representation of the internal organization of Regulatory Network Analysis Tool and the way in which queries are performed, the RESTful Web Service communicates the user query with the database and then the response with the web interface. At the bottom are examples of web service calls.

(color, shape, size, labels, etc.), searching elements, filtering genes and performing some simple exploratory analyses that include adding new genes to the network, including the network in the context of Reactome (7), upload a file with gene attributes that can further be used for filtering purposes and uploading gene expression values that can be also used for filtering or color-labeling the genes.

Depending on the data available, two types of tests are possible. If the list of genes represents a set of differentially expressed genes, a single enrichment analysis of regulatory elements over-representation is carried out (5). A conventional Fisher exact test is used to calculate the enrichment. If the input data are list of genes ranked by a value that accounts for differential expression (e.g. the statistic used) then a variant of the GSEA (6) is carried out. This last approach is much more sensitive than the conventional enrichment analysis. As many TFs are simultaneously checked for enrichment, multiple testing effect is corrected by the popular false discovery rate method (8).

It is worth mentioning that the results are conceptually different in the case of TF and miRNAs, given that the former are positive regulators and the latter are negative regulators. Thus, if we are analysing genes activated, then the results on TF make reference to the TFs active, whereas the results on miRNA would make reference to miRNAs inactive.

The web server can maintain sessions and save data and results obtained for new sessions.

Databases and motif finding

The information on transcription factor binding sites (TFBSs) and miRNA target sites has been collected from different databases. Information on TFBSs was taken from Ensembl (Release 64, GRCh37) (9) through the

corresponding application programming interface. In the case of miRNAs, we have used miRBase (release 18, November 2011) (10), the microrna.org (Release August 2010), the TargetScan (11) predictions, as well as more curated information from miRecords (release 3, November 2010) (12) and miRTarBase (release 2.5, October 15, 2011) (13). Information of diseases has been taken from The human miRNA disease database (release November 2011) (3) and Phenomir (14). At present, information for human and mouse is available. In coming versions, more species will be added.

MiRNA targets are predicted by searching for matches with the miRNA seed regions (11). In the case of TF, the use of a position-specific weight matrix (15) that accounts for the DNA binding preferences of the TF, is used for detection of TFBSs. These binding sites are mapped into the promoter region of every gene, establishing the connections between TFs and genes.

Other tools

Many single enrichment tools include enrichment analysis of TFBSs, such as the FatiGO+ (16), DAVID (17) and other (18). More recently, some tools arise specifically focused on the analysis of regulatory motifs in the genes of a list. Thus, tools like Pscan (19), TransFind (20), oPOSSUM (21) or CARRIE (22) carry out different variants of enrichment analysis and report lists of over-represented TFBSs alone or in combination (23), present in the set of genes analysed. Similar applications have been published for miRNAs, like MMIA (24), GeneSet2miRNA (25), miRvestigator (26), including the analysis of ensembles of miRNAs, such as MiRror (27). A more sophisticated tool, the mirConnX (28), combines miRNA and TFs and also provides a graphical interface based on Cytoscape web (29), which is a Flash application

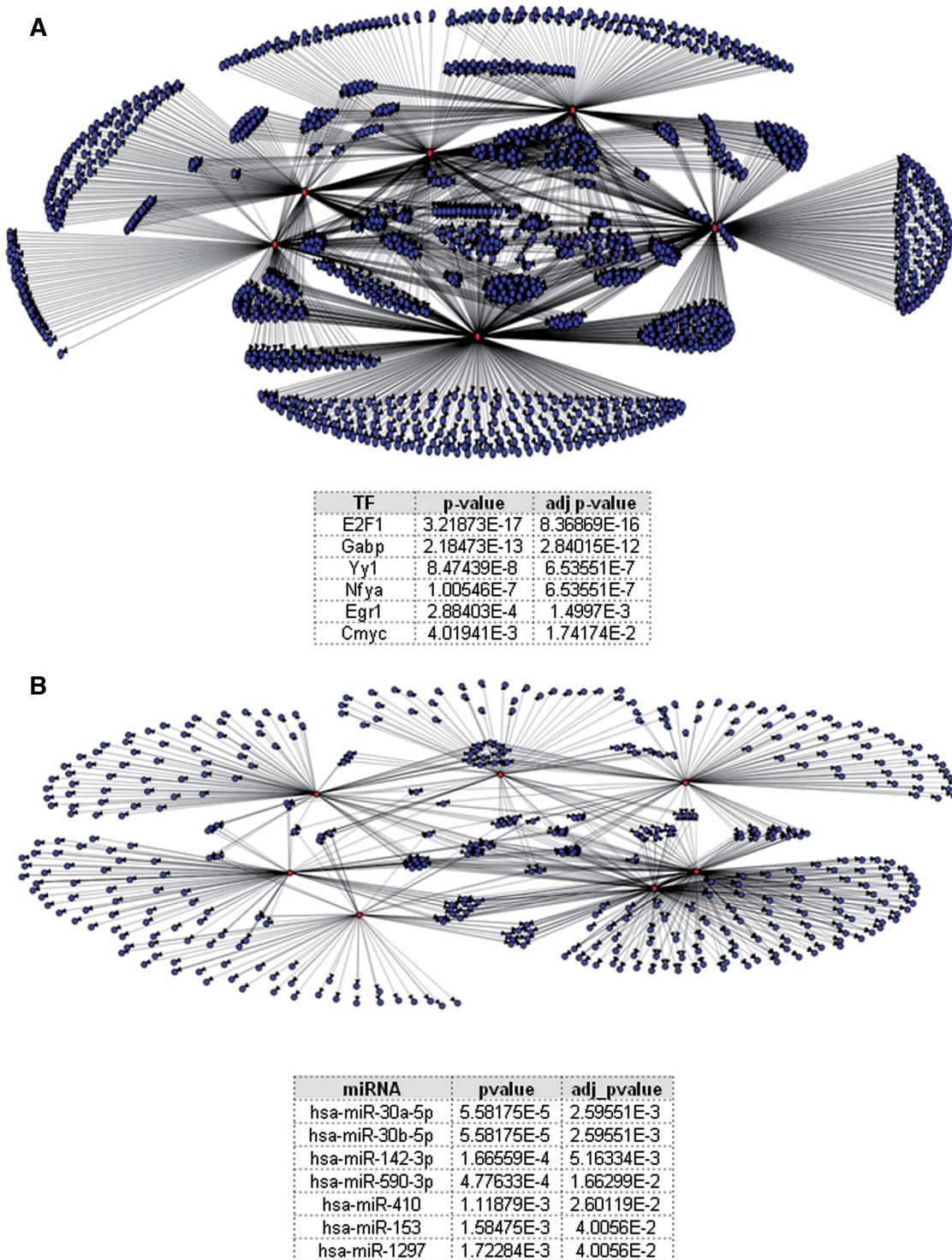


Figure 2. An example with Fanconi anemia. Significantly expressed genes in a comparison of Fanconi anemia patients with healthy controls have been analysed by enrichment analysis. (A) Transcription factors with an adjusted $P < 0.05$ have been represented in the network (in red), along with the target genes regulated by them. (B) miRNAs with an adjusted $P < 0.05$ have been represented in the network. As explained in the text, the transcription factors are compatible with the genes activated, whereas the miRNAs should be downregulated to be compatible with the genes activated.

that depends on the corresponding plug-in. RENATO, in addition to single enrichment analysis, offers the GSEA, which is known to be still more sensitive in finding enrichments (6,18,30). An advanced graphical interface is also

provided, based on modern and efficient HTML5, with much more options (e.g., information on the genes and different layouts) than other interfaces and much more optimized than a Flash application.

RESULTS

An example with Fanconi anemia

We illustrate this with an example of Fanconi anemia (FA), a disease in which signaling is relevant. FA is a rare chromosome instability syndrome characterized by aplastic anemia and cancer and leukemia susceptibility (31). A recent study uses gene expression microarrays to identify differences at the transcription level between bone marrow cells from normal volunteers and from children and adults with FA (32). Eleven normal volunteers and 21 patients were studied. Gene expression datasets for FA were retrieved from the GEO database. Differential expression analysis was carried out using the Babelomics suite (33). RENATO has been used to study the set of regulatory elements compatible with the observed gene deregulation. Figure 2 shows the regulatory network that justifies the observed increase in the gene activity caused by the disease. Six TFs (E2F1, Gabp, Yy1, Nfyα, Egr1, Cmyc) account for activation observed in the disease (Figure 2A). Some of these TFs have already been linked to FA (34). Similarly, the miRNAs that must be switched off to produce the observed gene expression levels are listed in Figure 2B, along with the regulatory miRNA network. Some experiments of inhibition of DNA repair link FANCG gene to several miRNAs, among which some of the detected in this analysis are included (hsa-miR-30a-5p) (35).

DISCUSSION

RENATO is a convenient interactive web server for the analysis gene expression profiling experiment that finds the regulators (TF or miRNA) that significantly account for the gene activity observed. The program also implements a highly efficient interactive interface, based on HTML5, with many options, and able to offer much information of the genes and regulators involved on the experiment analysed.

ACKNOWLEDGEMENTS

We also thank the support of the National Institute of Bioinformatics (www.inab.org) and the CIBER de Enfermedades Raras (CIBERER), both initiatives of the ISCIII, MINECO.

FUNDING

Spanish Ministry of Science and Innovation (MICINN) [BIO2011-27069]; Conselleria de Educacio of the Valencian Community [PROMETEO/2010/001]; Red Tematica de Investigacion Cooperativa en Cancer (RTICC), ISCIII [RD06/0020/1019, partial].

Conflict of interest statement. None declared.

REFERENCES

1. Hobert,O. (2008) Gene regulation by transcription factors and microRNAs. *Science*, **319**, 1785–1786.
2. Boyadjiev,S.A. and Jabs,E.W. (2000) Online Mendelian Inheritance in Man (OMIM) as a knowledgebase for human developmental disorders. *Clin Genet*, **57**, 253–266.
3. Lu,M., Zhang,Q., Deng,M., Miao,J., Guo,Y., Gao,W. and Cui,Q. (2008) An analysis of human microRNA and disease associations. *PLoS ONE*, **3**, e3420.
4. Furney,S.J., Higgins,D.G., Ouzounis,C.A. and Lopez-Bigas,N. (2006) Structural and functional properties of genes involved in human cancer. *BMC Genomics*, **7**, 3.
5. Al-Shahrour,F., Diaz-Uriarte,R. and Dopazo,J. (2004) Fatigo: a web tool for finding significant associations of gene ontology terms with groups of genes. *Bioinformatics*, **20**, 578–580.
6. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*, **102**, 15545–15550.
7. Croft,D., O'Kelly,G., Wu,G., Haw,R., Gillespie,M., Matthews,L., Caudy,M., Garapati,P., Gopinath,G., Jassal,B. et al. (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res*, **39**, D691–D697.
8. Benjamini,Y. and Yekutieli,D. (2001) The control of false discovery rate in multiple testing under dependency. *Ann Statist*, **29**, 1165–1188.
9. Flicek,P., Amode,M.R., Barrell,D., Beal,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fairley,S., Fitzgerald,S. et al. (2012) Ensembl 2012. *Nucleic Acids Res*, **40**, D84–D90.
10. Griffiths-Jones,S., Grocock,R.J., van Dongen,S., Bateman,A. and Enright,A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*, **34**, D140–D144.
11. Friedman,R.C., Farh,K.K., Burge,C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*, **19**, 92–105.
12. Xiao,F., Zuo,Z., Cai,G., Kang,S., Gao,X. and Li,T. (2009) miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res*, **37**, D105–D110.
13. Hsu,S.D., Lin,F.M., Wu,W.Y., Liang,C., Huang,W.C., Chan,W.L., Tsai,W.T., Chen,G.Z., Lee,C.J., Chiu,C.M. et al. (2011) miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res*, **39**, D163–D169.
14. Ruepp,A., Kowarsch,A. and Theis,F. (2012) PhenomiR: microRNAs in human diseases and biological processes. *Methods Mol Biol*, **822**, 249–260.
15. Stormo,G.D. (2000) DNA binding sites: representation and discovery. *Bioinformatics*, **16**, 16–23.
16. Al-Shahrour,F., Minguez,P., Tarraga,J., Medina,I., Alloza,E., Montaner,D. and Dopazo,J. (2007) Fatigo +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. *Nucleic Acids Res*, **35**, W91–W96.
17. Sherman,B.T., Huang da,W., Tan,Q., Guo,Y., Bour,S., Liu,D., Stephens,R., Baseler,M.W., Lane,H.C. and Lempicki,R.A. (2007) DAVID knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. *BMC Bioinformatics*, **8**, 426.
18. Dopazo,J. (2006) Functional interpretation of microarray experiments. *Omics*, **10**, 398–410.
19. Zambelli,F., Pesole,G. and Pavese,G. (2009) Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res*, **37**, W247–252.
20. Kielbasa,S.M., Klein,H., Roider,H.G., Vingron,M. and Bluthgen,N. (2010) TransFind—predicting transcriptional regulators for gene sets. *Nucleic Acids Res*, **38**, W275–W280.
21. Ho Sui,S.J., Fulton,D.L., Arenillas,D.J., Kwon,A.T. and Wasserman,W.W. (2007) oPOSSUM: integrated tools for analysis of regulatory motif over-representation. *Nucleic Acids Res*, **35**, W245–W252.
22. Haverty,P.M., Frith,M.C. and Weng,Z. (2004) CARRIE web service: automated transcriptional regulatory network inference and interactive analysis. *Nucleic Acids Res*, **32**, W213–W216.

23. Waleev,T., Shtokalo,D., Konovalova,T., Voss,N., Cheremushkin,E., Stegmaier,P., Kel-Margoulis,O., Wingender,E. and Kel,A. (2006) Composite module analyst: identification of transcription factor binding site combinations using genetic algorithm. *Nucleic Acids Res.*, **34**, W541–W545.
24. Nam,S., Li,M., Choi,K., Balch,C., Kim,S. and Nephew,K.P. (2009) MicroRNA and mRNA integrated analysis (MMIA): a web tool for examining biological functions of microRNA expression. *Nucleic Acids Res.*, **37**, W356–W362.
25. Antonov,A.V., Dietmann,S., Wong,P., Lutter,D. and Mewes,H.W. (2009) GeneSet2miRNA: finding the signature of cooperative miRNA activities in the gene lists. *Nucleic Acids Res.*, **37**, W323–W328.
26. Plaisier,C.L., Bare,J.C. and Baliga,N.S. (2011) miRvestigator: web application to identify miRNAs responsible for co-regulated gene expression patterns discovered through transcriptome profiling. *Nucleic Acids Res.*, **39**, W125–W131.
27. Friedman,Y., Naamati,G. and Linial,M. (2010) MiRrror: a combinatorial analysis web tool for ensembles of microRNAs and their targets. *Bioinformatics*, **26**, 1920–1921.
28. Huang,G.T., Athanassiou,C. and Benos,P.V. (2011) mirConnX: condition-specific mRNA-microRNA network integrator. *Nucleic Acids Res.*, **39**, W416–W423.
29. Lopes,C.T., Franz,M., Kazi,F., Donaldson,S.L., Morris,Q. and Bader,G.D. (2010) Cytoscape web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.
30. Huang da,W., Sherman,B.T. and Lempicki,R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.*, **37**, 1–13.
31. Taniguchi,T. and D'Andrea,A.D. (2006) Molecular pathogenesis of Fanconi anemia: recent progress. *Blood*, **107**, 4223–4233.
32. Vanderwerf,S.M., Svahn,J., Olson,S., Rathbun,R.K., Harrington,C., Yates,J., Keeble,W., Anderson,D.C., Anur,P., Pereira,N.F. et al. (2009) TLR8-dependent TNF-(alpha) overexpression in Fanconi anemia group C cells. *Blood*, **114**, 5290–5298.
33. Medina,I., Carbonell,J., Pulido,L., Madeira,S.C., Goetz,S., Conesa,A., Tarraga,J., Pascual-Montano,A., Nogales-Cadenas,R., Santoyo,J. et al. (2010) Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. *Nucleic Acids Res.*, **38**, W210–W213.
34. Meier,D. and Schindler,D. (2011) Fanconi anemia core complex gene promoters harbor conserved transcription regulatory elements. *PLoS ONE*, **6**, e22911.
35. Tsai,Y.S., Lin,C.S., Chiang,S.L., Lee,C.H., Lee,K.W. and Ko,Y.C. (2011) Areca nut induces miR-23a and inhibits repair of DNA double-strand breaks by targeting FANCG. *Toxicol Sci*, **123**, 480–490.